

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 1, 2001, 15:52:23 ; Search time 170.72 Seconds  
(without alignments)  
264.889 Million cell updates/sec

Title: US-09-331-631A-3

Perfect score: 3532  
Sequence: 1 MAINTSNLCSLFLFLSLFL.....SSRSTKQOQPLVSLDFVGF 666

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 195891 seqs, 67900655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: PIR66:\*  
2: PIR1:\*  
3: PIR3:\*  
4: PIR4:\*  
Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1191.5	33.7	810	2 T44430	protein PV100 [imp
2	1142	32.3	566	2 S22477	vicilin precursor
3	1092	30.9	605	2 S06398	alpha-globulin typ
4	1083.5	30.7	588	1 FMCNAB	alpha-globulin B p
5	1022.5	28.9	509	2 S08059	alpha-globulin typ
6	902	25.3	573	2 A53234	globulin-1S, GluB1S
7	895	25.3	582	2 B53234	vicilin-like stora
8	875	24.8	637	2 B53221	globulin Bgl prec
9	872.5	24.7	639	2 B24810	beta-conglycinin a
10	859.5	24.3	605	2 S20007	beta-conglycinin a
11	855.5	24.2	605	1 FMS1BA	beta-conglycinin a
12	832.5	23.6	571	2 S00566	convicillin precurs
13	798	22.6	524	2 JQ1730	62K sucrose-bindin
14	770.5	21.8	483	2 T06459	vicilin-2 precurs
15	765	21.7	448	1 S24756	vicilin-like stora
16	760.5	21.5	433	2 S00567	vicilin precursor
17	760.5	21.5	459	2 S08505	vicilin - garden p
18	749.5	21.2	438	2 S35757	vicilin, 47K - gat
19	748	21.2	439	1 FMSYCB	beta-conglycinin b
20	727	20.6	450	2 S15675	globulin-2 precurs
21	715.5	20.3	463	2 A27288	vicilin precursor
22	711.5	20.1	463	2 S06309	vicilin precursor
23	704.5	19.9	410	1 FMPWVB	vicilin B precurs
24	689.5	19.5	445	2 JQ2264	canavalin - jack b
25	682.5	19.3	445	2 S00281	canavalin - sword
26	568	16.1	436	2 A23498	phaseolin type alp
27	565	16.0	430	2 S10156	alpha-phaseolin pr
28	557	15.8	421	2 A24810	phaseolin beta cha
29	552.5	15.6	414	2 S60190	vicilin - Zamia fu

ALIGNMENTS

RESULT 1	protein PV100 [imported] - winter squash	
T44430	C:Species: Cucurbita maxima (winter squash)	
	C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #extl_change 21-Jul-2000	
	C:Accession: T44430	
R.Yamada, K.; Shimada, T.; Kondo, M.; Nishimura, M.; Hara-Nishimura, I.		
J. Biol. Chem. 274, 2563-2570, 1999		
A:Title: Multiple functional proteins are produced by cleaving Asn-Gln bonds of a sin		
A:Reference number: 222767; PMID:99107919		
A:Accession: T44430		
A:Status: preliminary; translated from GB/EMBL/DBJ		
A:Molecule type: mRNA		
A:Residues: 1-810<YAM>		
A:Cross-References: EMBL:AB019195; MID:g3808061; PIDN:BAA34056.1; PID:g3808062		
Query Match	33.7%; Score 1191.5; DB 2; Length 810;	
Best local similarity	34.3%; Pred. No. 1.6e-66;	
Matches 284; Conservative 144; Mismatches 218; Indels 183; Gaps 23;		
OY 1	MAINTSNLCSLFLFLSLFLSTVSLAE-----SEP- 31	vicilin-like stora
DB 1	MALSKVKLRLLAFTLFLACLISVGLDGKESLSGAGVDHOCVNRCELKGNVDEFA 60	vicilin A precursor
OY 32	-----DRQYECKRQCMQLETS-GQMRCVSCQCKREE----- 65	globulin-1 - maize
DB 61	ACKKGGGVNORGSPRAVEYECRLRCVAERGVDQKCEQVEERLRERBOGGEVDDEV 120	probable major pro
OY 66	---DIMSKY-----DNDPPQTDQOQORRCROE 93	globulin1 - maize
DB 121	ERRDPMEEREOQRREHEERERERERERERGRRRDENEDPRREDEORR-REDE 178	convicillin precurs
OY 94	SGPRQOQOYCRCKELCEEEVEYNRQDPQOQYEQCOERQHEHPRMQTCQRCERR 153	beta-conglycinin a
DB 179	RRRRRDEQERERERGRDEEDDEN-QRDPWMREQ-----ERREQRRRREDEQERREQ 232	vicilin-2 precurs
OY 154	Y-----EKE-----KKQOKRYEQOQREDEKLEEMKEDKRP-----QOREYE 195	vicilin-like stora
DB 233	RRCGRDEDEQNRDPDMRRQKRRQEOERERERERGRGRDDEQNRDWMREDEERRQ 292	vicilin precursor
OY 196	DCRRRCQOE-----PROOYQORRCRCROHONGGG-- 227	vicilin precursor
DB 293	ERRRRRDEQERERERGRGRDVEDENQRPDMRREDEERRRRREDEQERERERHGRRE 352	vicilin B precursor
OY 228	-----DLNPORGSGRYEE-----GEKQSDNPYVFERSLSFTRFEEGHISVL 273	canavalin - jack b
DB 353	EQSRDERRRHRHROGGSRVNQVAIRTEQDSNNPYFQGRQNSVRSDQGHWRVL 412	canavalin - sword
OY 274	ENYGGSKLLRLAKNRLVLEENPNAFVLPHTLDDALLVYGGGALKMTHRRN--- 330	phaseolin type alp
DB 413	ERSESRSEILKGIKNORLALLEARPHFTFVPHHLDAECVLLVVRGRATITTVQOEKRTR 472	alpha-phaseolin pr

```

OY 331 -ESNLECCGVIRIPATTFYLLINBONNENHLAKFLOTISTSTGOXKKEPPGAGGONDEY 389
Db 473 KESTNVSQGDWITPATTLTYLNOE-NEDLOLVKIVOPVNNNGEKKDYLSAGES-QAY 530
OY 390 LSTPKEKLEALNTOAERLGVAGGOORE--GVIIASOEOIRELRDSESRMHIRRG 447
Db 531 YSVNSVDLEALINIPDKLERIFKQRRREGKIVASOEBALSO-----RATSVRG 585
OY 448 GESSRGPYNLFNKRPILSNKYGOAYEVKPEPDYBOLODMDSVFIANTITGSMNGPEFNTR 507
Db 586 SRGVRAPIKLESQPPYNNQYGMFEACPDPEFQLRTDVATSVDIKGGMMVPIHNSR 645
OY 508 STKYVYVASEADYENACPHLS-----GHGRGRCGKKRHEEE---DVHYEQVAKRLS 557
Db 646 ATWVVFSEGAGSEFMACPHIOSSOWOMGRHEERHEMRKREEREREBSRGREFRYAGRLS 705
OY 558 KRELIYVPGHPYVFEVSSGNEMLLFAFGINAGNNHNFPLAGERVYLOOIEPOAELAF 617
Db 706 EGVLYVTPACHPIAIMASPENIRLVGFIAGNANHNFPLAGRE-NIMNELDREAKELAF 764
OY 618 AARPEVEELLNSODESIFPCPRHOHOGSSRSTKOOPJVLSTLDVEGF 666
Db 765 NVECKQADEIFRSORESFTEGPEGGGRNST---ERSPLSTILKLAGI 809

```

RESULT	2
S22477	
vicillin precursor - cacao	
C:Species:	Theobroma cacao (cacao)
C:Date:	30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 17-Mar-1999
C:Accession:	S22477; S22478; S18105; S22050
R:McHenry, L.; Fritz, P.J.	
Plant Mol. Biol.	18, 1173-1176, 1992
A:Reference number:	S22477; M01D:92286309
A:Accession:	S22477
A:Molecule type:	DNA
A:Residues:	1-566 <MC2>
A:Cross-references:	EMBL:X62625
A:Accession:	S22478
A:Molecule type:	mRNA
A:Residues:	1-452 <MC2>
A:Cross-references:	EMBL:X62626
C:Genetics:	
A:Introns:	211/1; 269/3; 296/3; 391/3; 502/1
C:Superfamily:	glycinin
C:Keywords:	seed; storage protein
F:1-24/Domain:	signal sequence #status predicted <SIG>
F:25-566/Product:	vicillin #status predicted <MAT>

[illegible][illegible]

```

RESULT      3
S063398
alpha-globulin type A precursor - upland cotton
N:Alternate names: seed storage protein
C:Species: Gossypium hirsutum (upland cotton)
C:Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 30-Sep-1993
C:Accession: S063398
P:Chlan, C.A.; Borroto, K.; Kamalay, J.A.; Dure III, L.
Plant Mol. Biol. 9, 533-546, 1987
A:Title: Developmental biochemistry of cottonseed embryogenesis and germination. XIX
A:Reference number: S063398
A:Accession: S063398
A>Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-605 <CHL>
C:Superfamily: glycinin
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-605/Product: alpha-globulin type A #status predicted <MAT>

```

Query Match Similarity	30.9%	Score 1092;	DB 2;	Length 605;
Best Local Similarity	36.2%	Pred. No. 1.7e-60;		
Matches 249;	Conservative 118;	Mismatches 210;	Indels 110;	Gaps 17.
QY	1	MAINTNLSLFLSLFLSTVSLAE-----SEPD-ROEYBECKROCMOLETSQOMK	53	
Db	1	MAVNSVAVVLLLF--SLFLSFLGLCSAKDPGRSDEDDPQQRYEDCKKRC-QLETSQGE	57	
QY	54	RCVSGCDKRFREEDIDMSKYDNDDDPQTQCCQQCRRCROESGSPRQOQYCSRRCKTECEE	113	
Db	58	Q--DKCEDRSSEQLKEEQRQDGEDPQRYQDRCRHCQOQEEERRLRP-----HCQSCREQ	109	
QY	114	EEYNRRDRDQOQYEOQSORQENRTEPRNMHTQCSORCERAKERKRYKQRYEQQREDE	173	
Db	110	YEQOQOQOQDRQKFEQQRQCSQMOQERYRKRQOCSYKCEQYQOEPRMKGER-----E	160	
QY	174	EKYEEKKEEDNKRDPQOQREYEDCRRCRQEOERPOOYQOSCRKCEQROHQRGGDLINPQ	233	
Db	161	NKMREREEDES-----	171	
QY	234	RQSGSGRTEGGEQKOSNPRTYFDEKSLSTRFTREBCHLSYLENRYGRSKLLRALKNTRYLV	293	
Db	172	-----DEGEQOQRNPRTYFHRRSFQRFREEGNFRVLDQREFADKNNLLRGNGIERIAI	224	
QY	294	LEANPAFLPRLNLADALILYTGGRGSKALMINIDRNSVNLBEGDVIIRIPACTYVLIIN	353	
Db	225	LEANPRTFVLPRLHCSAEKIYYVTNRGRGIVLTVTYHNKESINAVPGVVYRIPACSTVYLIAN	284	
QY	354	RDNRRRLTAFLKFLQITSTPQGYKEFPFAGGQGNRPYLSFESKELLEALNTQALRGVLI	413	

[illegible]

RESULT 4

FWCNCAB

alpha-globulin B precursor (clone C72) - upland cotton

N/Alternate names: seed storage protein; vicilin precursor

C/Species: *Gossypium hirsutum* (upland cotton)

C/Date: 30-Sep-1991 #sequence\_revision 30-Sep-1991 #text\_change 16-Jul-1999

C/Accession: A30838; 506911

R/Chan, C.A.; Pyle, J.B.; Legocki, A.B.; Dure III, L.

Plant Mol. Biol. 7, 475-489, 1986

A/Title: Developmental biochemistry of cottonseed embryogenesis and germination XVIII

A/Reference number: A30838

A/Accession: A30838

A/Molecule type: mRNA

A/Residues: 1-588 <CH2>

A/Cross-references: GB:M16891; NID:g167374; PIDN:AAA33071.1; PID:g167375

A/Experimental sources: var. Coker 201

R/Chan, C.A.; Borroto, K.; Kamalay, J.A.; Dure III, L.

Plant Mol. Biol. 9, 533-546, 1987

A/Title: Developmental biochemistry of cottonseed embryogenesis and germination. XIX. Seed

A/Reference number: 506398

A/Accession: 506911

A/Status: not compared with conceptual translation

A/Molecule type: DNA

A/Residues: 1-81 <CH2>

C/Comment: This is a seed storage protein.

C/Superfamily: glycinin

C/Keywords: glycoprotein; seed; storage protein

F/1-25/Domain: signal sequence #status predicted <IG>

F/26-588/Product: alpha-globulin storage protein #status predicted <MAT>

F/417/Binding site: carbohydrate (Asn) (covalent) #status predicted

	Query Match	Similarity	38.7%	Score 1083.5	DB 1	Length 588	
	Best Local	Similarity	38.2%	Pred. No. 5.5e-60			
	Matches 229	Conservative 114	Mismatches 182	Indels 75	Gaps 13		
QY	66	DDPOTCCOCCORCROEESGPRPOOYOCORCKEICEEEEEYROR--PPOOYEOCOERC	133				
Db	35	DDPRTYEDCRRRCRMDTRGCKEEOOOCCEESCKSYQGEKDQRRHRPEDPQRTYECCQEC	94				
QY	134	ORHETEPHMOITCOORCERRTYEKKRKKOAKRYEEOORDEEKEYEERMKEEDNKRDPQORE	193				
Db	95	R--OOEORORPOCCORCLKREOEEOOQ-----SQO	123				
QY	194	YEDCRRREOEE--PROOYOCORCREEOORHGREGDILNPR-----GSGRYEEBEE	245				
Db	124	FOECOOHCODQORPRTKQOCCVRECRERYOE-----NPRRRREDEABEEETEEBQO	175				
QY	246	KSDNPYFDEKSLSTRTEEGHISYLENFGRSKLRLAKNRYRLVLEAPNPAFLPT	305				

[illegible]

RESULT 5  
S08059  
alpha-globulin type B precursor (tandem 1) - upland cotton (fragment)  
N:Alternate names: seed storage protein  
C:Species: *Gossypium hirsutum* (upland cotton)  
C:Date: 31-Mar-1990 #sequence\_revision 31-Mar-1990 #text\_change 30-Sep-1993  
C:Accession: S08059  
R:Chuan, C.A.; Borroto, K.; Kamalay, J.A.; Dure III, L.  
Plant Mol. Biol. 9, 533-546, 1987  
A:Title: Developmental biochemistry of cottonseed embryogenesis and germination. XIX  
A:Reference number: S06398  
A:Accession: S08059  
A:Status: not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-509 <CHL>  
C:Superfamily: glycinin

	Query Match	28.9%	Score 1022.5	DB 2	Length 509
	Best Local Similarity	30.8%	Pred. No. 2.8e-56		
	Matches 217	Conservative 105	Mismatches 153	Indels 85	Gaps 13
QY	121	DPDOYEECCORSCOHNEVPRHMTQCOORCERFRYKFKRKQOKRYEEDQREDEKYEEDRM	180		
DB	1	DPQRYEECCQDECR--QGEERQDPOCQDRLKRFQEBQDQ-----	38		
QY	181	KEEDNKRDPQOQREYEDCRRRCQOE--PROOYQORRCRODROHGRGDLINPORG---	235		
DB	39	-----SQROQECQCHQDQOQREKQOQCVCREKRYQE-----NPRGGERE	81		
QY	236	---GSGREEEEEEKOSNDPYTFDERSJLSTRRTBEGHLSVLNENFYGRSKLLPALKNYRLV	292		
DB	82	EEAEEETEETEEEOESQHNPFHFHRSFSQSRFEHGNFRVQRFASRHPILRGJIEPFLS	141		
QY	293	LLENPNPAFVLPRLHDAVALILVGGRCALMLHRDNEESNLECGDVIRLPACTFTPLI	352		
DB	142	ILEANPNLFEVPHHCDAKITYLVNIGKSTLFLFHEMKESTNIVPGVYKVPAGSTVYLA	201		
QY	353	NRDNNERLIATKLFQTIQSTPGQYKEFPFAGQONEPPLYSTFSKEILLALMTQAEIRLGV	412		
DB	202	NQDKKEKLIIVLHRPVNPNQGFEEFPFAGSQOROSTYLRAFSRILFEAFMTRSQDEL	261		
QY	413	LG-----QORECVIISASQDILRELTRDSESRHMIIRGCESSRCPVMLPNKRPLYS	465		
DB	262	FGGRQSRRRQDQGMFRKASQDQIRALSOEATSR---EKSGE--RAFAPLILSTQTPYS	315		









Db 491 LIGRLYLPHKDKESFPFPELREDEGRRA 522

# RESULT 14

T06459

62k sucrose-binding protein homolog - garden pea

N:Alternate names: P54 protein

C:Species: Pisum sativum (garden pea)

C>Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 20-Jun-2000

C:Accession: T06459

R:Rodrigo II, M.I.

Submitted to the EMBL Data Library, February 1997

A:Reference number: Z15693

A:Accession: T06459

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-483 <R0D>

A:Cross-references: EMBL:Y11207; PIDN:CAA72090.1

A:Experimental source: var. Lincoln

C:genetics:

A:Note: p54

C:Superfamily: glycinin

Query Match 21.8%; Score 770.5; DB 2; Length 483;

Best Local Similarity 31.6%; Pred. No. 1.2e-40;

Matches 163; Conservative 113; Mismatches 169; Indels 71; Gaps 12;

QY 135 RHTEPRHMQTCGRCERREKREKQRYEQREDEKYEERKKEDNKRDPQORE 194

Db 28 RKEKDP-ELTTCDDQCD-----MOKYDEDK----- 53

QY 195 EDCRRCEQDEPRQOQOCRCRCQOORGRGDLINPORGSGRYEERKESQDNPPYF 254

Db 54 RIMERCSD-----DYIKKKQERQKKHEH-----EEBEQEQEEDENPYF 93

QY 255 DEBSLSTRTTEEGHISVLENFYGRSKLLRALKNYRLVLEANNPVLFTLHDADAILL 314

Db 94 EDDPFETKIDTKGRVLLINKFNESKSLKLNIEYGLAVLEIKANAFILSPHHYDSEALIF 153

QY 315 VTGRCALMKIHNDRNRESYNLECGDVIRIPAGTFYLINNDNERLHIAKF--LQTI 371

Db 154 NIKGRGIIIGLVADRTERENLEEGDITKRPAGTPTLVNDENEKLYIAFHMPPSSGSA 213

QY 372 PGQYKEFFPAGGONPEPYLSTFESKILEALNTOAERLQGOQREGVLIISASQOIRF 431

Db 214 PVNLPEFPESAGKRPESVLTNTFSYVLQALAKSKSGELETVLDEQKKGRIFKIEKEDVRC 273

QY 432 LTRDDSESRMHIRRGESRGPYNLFNKRPLYSNKYGOAYEVKPEDYRQ-LQDMVSVF 490

Db 274 LAPKKS--LWPF--GDPFKSPFNIFSNNPAFSNKFGLFEVGPSQEKSGLEGLNIMLT 327

QY 491 IANITGSMWMPFNTRSRKVVVAVASGEADVACPHLSGRHGRGGRKHEEEDVHYE 550

Db 328 LANITGSMSTHYNTNANKIALYIDGEGELEMACPHMPPSSSSNSR-----QKSSISYH 382

QY 551 OVKAARLSKREAVVPYGHVYFVSGNENILLFAGFINAQNHNHNLVLAGERNVLTQIEP 610

Db 383 NINAKRLPGVMFVYPAGHFPVNIASKKNLIYVCFEVNAQNRKKLLALACK-KNIVSALDK 441

QY 611 QAMELAFAPRKEVELEFNSODESIFPGPROHQO 646

Db 442 AAKEVAFDIAAEKVDVEFERKEE-FFPYDNEERKE 476

# RESULT 15

S24756

vicilin-like storage protein precursor - white spruce

C:Species: Picea glauca (white spruce)

C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999

C:Accession: S24756; S18873

R:Newton, C.H.; Flinn, B.S.; Sutton, B.C.S.

Plant Mol. Biol. 20, 315-322, 1992

A>Title: Vicilin-like seed storage proteins in the gymnosperm interior spruce (Picea

A:Reference number: S24756; MUID:93004485

A:Accession: S24756

A:Molecule type: mRNA

A:Residues: 1-448 <NEW>

A:Cross-references: EMBL:X63191; NID:g20500; PIDN:CAA44873.1; PID:g20501

C:Superfamily: vicilin

Query Match 21.7%; Score 765; DB 1; Length 448;

Best Local Similarity 37.8%; Pred No. 2.4e-40;

Matches 162; Conservative 89; Mismatches 154; Indels 24; Gaps 9;

QY 232 PORGSGRYEERKESQDNPPYFDESLSTRTTEEGHISVLENFYGRSKLLRALKNYRL 291

Db 34 PEYLGRGRGR-EEBEENPNYVHSDSPFRASSAEGELRALPNFGEVSLLEGLCKIFV 92

QY 292 VLEANNPNAFVLTHTLHDADAILLVTTGRCALMKIHNDRNRESYNLECGDVIRIPAGTFY 351

Db 93 TCIEMKPNTVYMLPHYIDATWILVTYTRGRGVIAVHONELVKRLLEGDVFGVPSGHTFYL 152

QY 352 INEDNERLHIAKFLOTISF-PGQYKEFFPAGGONPEPYLSTFESKILEALNTOAERL 410

Db 153 VNDDNITRLIASLVPRVSTVGEIOPFYVAGRNPOITYSAFSDVLEAENVTNVOOLE 212

QY 411 GVLGQREGVLIISASQOIRELTTRDDSESRMHIRRG---GESSR---GPYNLFNKRPL 463

Db 213 RIRGKHSQVLIHANDEQIREMR-----KRGFSAGMSAPENHPKPFNLNRQKRD 262

QY 464 YSNKYQAYEVKEDYRQLODMVSVFIANITGSMWMPFNTRSRKVVVAVASGEADVEM 523

Db 263 FENENRFTIAGKKNYPFLDALDVSGLDLNPGSMTAPLSKSTISIGIVNGEGRIEM 322

QY 524 ACPHLSGRHGRGGRKHEEEDVHYEYVKARLSKREAVVPYGHVYFVSGNENILLF 583

Db 323 ACPHL-GQHGW--SSPREGDDITTYRWAKLRITGSYIVPAGHPIYETIASINSRLQIL 379

QY 584 AFGINANNHNLVLAGERNVLTQIEPQAMELAFAPR-KEYEELFNSODESIFPGPRO 642

Db 380 WFDLNRGNBRQGLACK-NNVLNLTLEIRIQLSTFNVPGRBEIEVLQAKQDQYILNGPQR 438

QY 643 HQOQSSRST 651

Db 439 RSRDEARSS 447

Search completed: March 1, 2001, 15:52:28  
Job time: 352 sec